Physica 27D (1987) 1–20
North-Holland, Amsterdam

RESONANCES IN AREA-PRESERVING MAPS

R.S. MACKAY[1], J.D. MEISS[2] and I.C. PERCIVAL[3]
[1]*Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK*
[2]*Institute for Fusion Studies, University of Texas, Austin, TX 78712, USA*
[3]*School of Mathematics, Queen Mary College, University of London, London E1 4NS, UK*

Received 26 June 1986
Revised manuscript received 19 January 1987

A resonance for an area-preserving map is a region of phase space delineated by "partial separatrices", curves formed from pieces of the stable and unstable manifold of hyperbolic periodic points. Each resonance has a central periodic orbit, which may be elliptic or hyperbolic with reflection. The partial separatrices have turnstiles like the partial barriers formed from cantori. In this paper we show that the areas of the resonances, as well as the turnstile areas, can be obtained from the actions of homoclinic orbits. Numerical results on the scaling of areas of resonances with period and parameter are given. Computations show that the resonances completely fill phase space when there are no invariant circles. Indeed, we prove that the collection of all hyperbolic cantori together with their partial barriers occupies zero area.

1. Introduction

In an integrable Hamiltonian system, most of the motion is quasiperiodic and confined to invariant tori. The simplest class of Hamiltonian systems for which the dynamics can be more complicated is that of area-preserving maps. In these systems the tori are one-dimensional, so they can be called invariant circles or invariant curves. For non-integrable maps there may exist some invariant curves; but many of the invariant curves are replaced by invariant Cantor sets [2–4], which we call "cantori".

In a previous paper [1] we initiated a theory of transport in area-preserving maps. Transport in chaotic regions is impeded by partial barriers which are formed on the framework of the cantori. Flux across these partial barriers takes place through "turnstiles". We believe that the partial barriers can be chosen so that the turnstiles line up in one or more "chimneys". Outside the chimneys there is no flux across the partial barriers. Choosing a discrete set of partial barriers (corresponding to the "most important" cantori) partitions the phase space into regions between which flux occurs, the amount being calculable as the difference in action of certain orbits. This allows one to estimate transport rates.

In this paper we continue the discussion of transport, providing a less arbitrary partition of phase space. The corresponding partial barriers are formed from pieces of stable and unstable manifold of hyperbolic periodic points, and the regions they bound we call "resonances".

Resonances are easily defined for integrable systems. Here we use "integrable" in the weak sense, permitting the integrals to have places where their derivatives fail to be independent. Then they can have isolated hyperbolic periodic orbits, with separatrices joining their points (consider for example, the time-1 map of the simple pendulum). The region interior to a separatrix is a resonance; it contains within it an orbit of the same period as the hyperbolic orbit. Each resonance is made up of a chain of such "islands". The number of islands is a multiple of the period of the central

0167-2789/87/$03.50

periodic orbit, the multiple commonly being unity.

Similar resonances appear in profusion when integrable systems are perturbed. However, the resonances do not appear to be well-defined, since narrow bands of apparently chaotic motion generically replace the well-defined separatrices. As the perturbation amplitude is increased, the sizes of two such resonances, as calculated by perturbation theory, can become large enough so that they would overlap, suggesting that all the invariant tori between the resonances are destroyed and there is "global chaos". This is the basis of the Chirikov resonance overlap method [5, 6].

In section 2 we recall a precise construction for resonances of non-integrable area-preserving maps [7, 8]. These resonances are bounded by "partial separatrices", formed from stable and unstable manifolds of hyperbolic periodic points. They have turnstiles just like the partial barriers formed from cantori. The flux in and out of resonances takes place only through these turnstiles. The area entering or leaving a resonance from above or below is related to the difference of action between pairs of heteroclinic orbits [1].

In section 3 we recall this flux formula, and extend it to obtain the area of a resonance. A resonance has a well-defined area regardless of whether its central periodic orbit is elliptic or hyperbolic with reflection, though not all this area may be accessible to a chaotic region. We will call the resonance elliptic or hyperbolic accordingly. We also find a formula for the area under the partial barrier formed from a cantorus.

As Birkhoff proved [9], each elliptic periodic point generically has elliptic periodic orbits which encircle it, and elliptic periodic orbits which encircle them, *ad infinitum*. This complicates matters, but we avoid this complication by restricting explicit calculations in this paper to resonances of the primary "class" [10], for which the central periodic orbits are rotational rather than vibrational. Our formulae are equally valid, however, for any class of resonance. We believe that with an appropriate convention, the regions occupied by resonances of the same class never overlap.

In section 4 we use these formulae to calculate some resonance areas. We also present numerical evidence which suggests that almost all of phase space is made up of resonances and invariant tori, the remaining parts, including all the cantori and their partial barriers, being of measure zero. We also present a proof, based on Poincaré recurrence, that the partial barriers formed from cantori have measure zero, provided they are hyperbolic.

We conclude in section 5 with a discussion of transport and of unresolved problems.

2. Resonances and chimneys

Suppose T is an area-preserving map. We denote phase space points by capital letters and their components by small letters, e.g. $X = (x, p)$. Orbits are denoted by subscripted variables, e.g. $X_{t+1} = TX_t$. The phase space is assumed to be a cylinder or annulus, with the configuration or angle variable, x, of period unity. The momentum coordinate is p. An example we shall use frequently is the "standard map":

$$p_{t+1} = p_t - k/2\pi \sin(2\pi x_t),$$
$$x_{t+1} = x_t + p_{t+1},$$

which has a single parameter k.

2.1. *Resonances*

We introduce the general theory of resonances with a particular example from the standard map. The 1/3 resonance is shown in fig. 1 for $k = 1.672$. The angle coordinates x_t of the period 3 orbits of this resonance are stationary points (x_0, x_1, x_2) of the action sum

$$W = F(x_0, x_1) + F(x_1, x_2) + F(x_2, x_0 + 1), \tag{2.1}$$

where the generating function F for two succes-

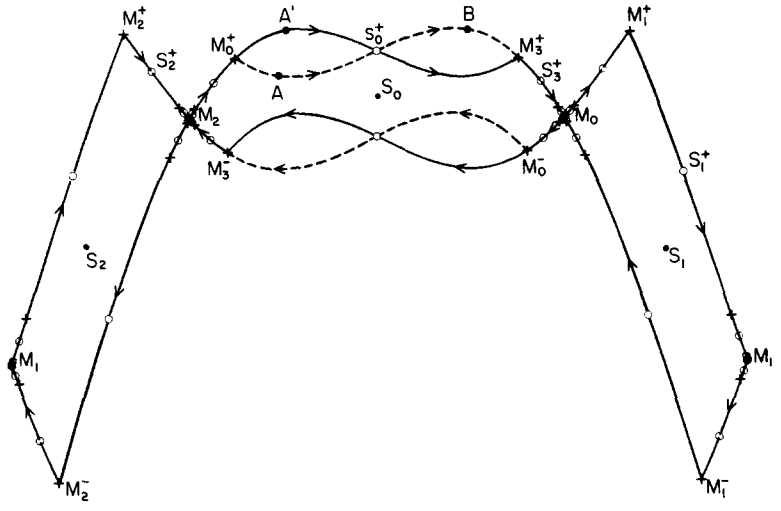

Fig. 1. Resonance of frequency 1/3 for the standard map in symmetry coordinates. M_t are points on the period 3 hyperbolic orbit, S_t are points on the period 3 elliptic orbit. Orbits homoclinic to M_t are labeled by "+" on the upper separatrix and "−" on the lower separatrix. The resonance boundary is indicated by solid lines. Dashed lines represent the other boundaries forming the turnstiles in the partial separatrices.

sive points of an orbit is

$$F(x, x') = \tfrac{1}{2}(x - x')^2 - V(x) \tag{2.2a}$$

with the potential

$$V(x) = -\frac{k}{(2\pi)^2}\cos(2\pi x). \tag{2.2b}$$

Stationarity of W gives a second difference form of the map T. The Hamiltonian form displayed above is obtained by defining the momentum as

$$p' = \frac{\partial F(x, x')}{\partial x'}. \tag{2.3a}$$

Stationarity of the action then implies

$$p = -\frac{\partial F(x, x')}{\partial x}. \tag{2.3b}$$

Any area-preserving map can be written in the form (2.3) for some generating function F, provided the twist condition $\partial x'/\partial p \neq 0$ is satisfied everywhere.

For $k \neq 0$ there are at least two stationary points of (2.1), a minimum and a "minimax." The points $\{M_t\}$ are the points of the periodic orbit of period 3 for which the action sum is a minimum. This orbit is normally an unstable hyperbolic orbit, but could exceptionally be parabolic (e.g. for $k = 0$). We will always assume that the minimising periodic orbits are hyperbolic. The points $\{S_t\}$ represent a periodic orbit of the same period for which the action sum is a minimax, i.e. a saddle point with only one downward direction in the space of the angle coordinates $(x_0, x_1, \ldots x_n = x_0 + m)$ of the orbit ($m = 1$, $n = 3$ in this case). This orbit may be a stable elliptic orbit or an unstable orbit that is hyperbolic with reflection: for present purposes it does not matter which.

For an integrable system there would be a separatrix joining the points M_t of any hyperbolic minimising orbit, to form a chain of islands; each island surrounds a point S_t of the minimax orbit. But for non-integrable systems there is only a "partial separatrix", whose structure is a bit more complicated. It is made up of pieces of stable and unstable manifolds of the minimising periodic points.

There is an upper partial separatrix and a lower partial separatrix, labelled in fig. 1 by superscript + and −, respectively. The solid lines represent the boundary of the 1/3 resonance, and the arrows represent asymptotic behaviour.

The upper partial separatrix is constructed from the upper unstable and stable manifolds of M_t. These manifolds must intersect. Choose any such point of intersection, say M_3^+ in fig. 1. The choice is arbitrary (see discussion below). The orbit of such an intersection point, shown as the +'s in the figure, is *homoclinic* to the orbit of M_0: it approaches this periodic orbit both in the past and in the figure. To construct the upper resonance boundary follow the right-going branch of the unstable manifold of M_2 until it reaches M_3^+. At M_3^+ switch to the stable manifold, and follow it to M_0. This forms a segment of the upper partial separatrix; the remaining segments are formed from its preimages. The first preimage connects M_1 to M_2 and the second M_0 to M_1. Thus after two preimages we obtain the upper solid curve shown in fig. 1, which is the upper resonance boundary.

The third preimage is back in the original gap, and consists of a longer piece of stable manifold, the extra piece appearing dotted in fig. 1, and a shorter piece of unstable manifold, joining at M_0^+. Area preservation implies there is an intersection point of the stable and unstable manifolds, S_0^+, between M_0^+ and M_3^+. The figure-of-eight formed by the stable and unstable segments between M_0^+ and M_3^+ is called the *turnstile*. The left lobe of the turnstile is the set of points which will cross the partial separatrix from below to above on the next iteration. Similarly, the right lobe is the set of points which will cross from above to below on the next iteration.

The lower partial separatrix has similar construction and properties, using lower stable and unstable manifolds. The resonance is the region bounded by the upper and lower separatrices. In fig. 1 it consists of the three curvilinear "rectangles". They all have equal area. In order to leave

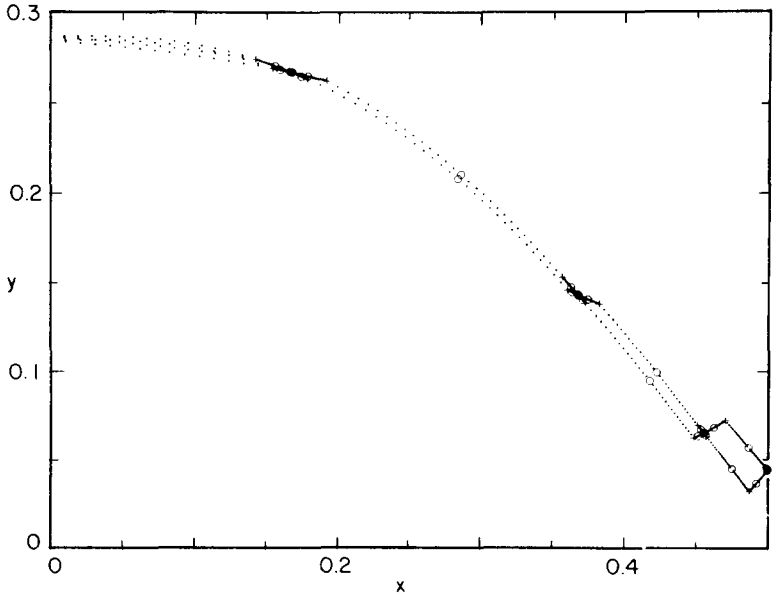

Fig. 2. Resonance of frequency 1/7 for the standard map at $k = 0.972$. The resonance is symmetric about $x = 0$, so only the $x > 0$ half is shown. The upper and lower separatrices are formed by the constrained minima of the action for 30 points in the largest gap around $x = 0$.

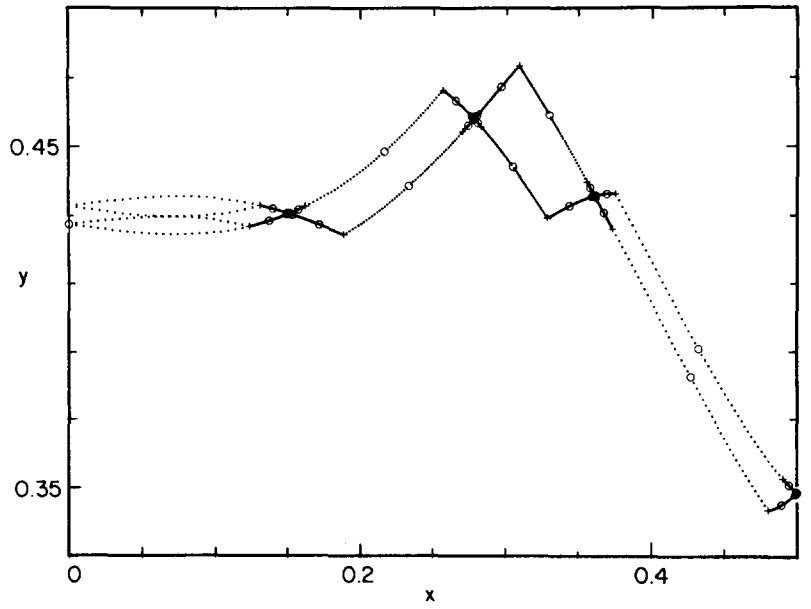

Fig. 3. Resonance of frequency 3/7 for the standard map at $k = 1.272$ in symmetry coordinates. The iterate of the island surrounding $x = 0$ is the rightmost island, surrounding $x = 0.43$.

the resonance an orbit must land in the left lobe of the upper turnstile or the right lobe of the lower turnstile.

In figs. 2 and 3 we illustrate the 1/7 and 3/7 resonances. For the second case, the apparently complicated vertical structure is due to big differences in momenta for islands that are neighbouring in configuration space but not in time.

Now we summarise the general construction of resonances for area-preserving twist maps.

(i) For all integers m, n with $n > 0$, there is a minimising m/n orbit, i.e. a minimum of $W_{m/n} = \sum_{t=0}^{n-1} F(x_t, x_{t+1})$ with $x_n = x_0 + m$.

(ii) Minimising periodic points are generically hyperbolic, so have stable and unstable manifolds, smooth curves W^s and W^u such that W^s is the set of points whose orbits are asymptotic to the orbit of the given point in the future, and similarly for W^u in the past.

(iii) The set of points belonging to minimising m/n orbits is monotone, i.e. the map preserves the angular order on the set. This allows one to speak of the gaps as coming in orbits.

(iv) In each orbit of gaps in the set of minimising m/n points, there is a minimax m/n orbit, i.e. saddle point of $W_{m/n}$.

(v) In each orbit of gaps in the set of minimising m/n points, there is a minimising orbit asymptotic to the orbit of the left end point in the past and to the orbit of the right end point in the future. We call this an m/n_+ heteroclinic orbit. Here "minimising" means that every finite segment of the orbit has minimum action with respect to variations fixing the ends. Similarly there is an m/n_- heteroclinic orbit asymptotic to the orbit of the right end in the past and to the orbit of the left end in the future.

(vi) Given an m/n_+ heteroclinic point, the right-going branch of the unstable manifold of the left endpoint of the gap and the right-going branch of the stable manifold of the right end intersect at that point. Similarly for m/n_- heteroclinic points, with right and left interchanged.

(vii) An upper partial separatrix can be formed by choosing one minimising m/n_+ heteroclinic point for each orbit of gaps in the set of minimising m/n points. Connect it to the left end point of

the gap with its unstable manifold and to the right end with its stable manifold, and connect the endpoints of the image gaps with the image of this. Similarly for lower partial separatrices.

(viii) The union of the sets of points of m/n orbits and m/n_+ orbits is monotone. Similarly for the union of m/n and m/n_- orbits.

(ix) For each orbit of gaps in the above there is a minimax m/n_+ (respectively m/n_-) heteroclinic orbit.

(x) The minimax heteroclinic points also lie on the corresponding stable and unstable manifolds and form the turnstile centres.

The arbitrariness in choice of minimising heteroclinic point affects the shape of the partial separatrices, but not the areas of the resonances, nor of the turnstiles.

There are also rotational partial barriers corresponding to each irrational number, formed on a framework of the cantorus and its minimax homoclinic orbit. While there is only one barrier for each irrational, the partial separatrices give two rotational partial barriers for each rational. These are in one-one correspondence with the continued fraction expansion for the reals: the upper partial barrier corresponds to the limit to a rational from above, and the lower partial barrier to the limit from below, e.g. since $1/3 = [0,2,1] = [0,3]$ we have

$$1/3|_+ \equiv \lim_{n\to\infty} [0,2,1,n],$$

$$1/3|_- \equiv \lim_{n\to\infty} [0,3,n].$$

For each rational continued fraction there are two periodic orbits: the minimising and minimax orbits. Similarly a limit of minimising (minimax) orbits approaches a minimising (minimax) orbit. These limits can be used as a guide to the computation of the partial barriers, as discussed in section 4.2 below.

Stable and unstable manifolds can be derived for hyperbolic minimising orbits from the action principle, and we believe that this will extend to the non-hyperbolic case. We discuss this derivation in appendix 1 for those interested.

2.2. *Chimneys*

There is a certain amount of freedom in defining the resonance boundary, corresponding to the choice of homoclinic points at which to switch from unstable to stable manifold in forming the partial separatrix. This does not matter too much when dealing with a single resonance. Indeed we shall see in section 3 that the area of a resonance and the areas of the upper and lower turnstiles are independent of these choices. However, the choice is important when it comes to fitting resonances of different rotation number together. We would like to make a choice such that resonances never overlap, as we aim to get a partition of phase space.

The standard map is particularly simple, as there is a natural choice. The cosine potential term $-V(x)$ in the generating function (2.2) has a maximum at the origin, so it seems reasonable that all orbits of absolute minimum action must avoid $x = 0$. Indeed, numerical experiment confirms that the configurations x_t of the minimising orbits all have their largest gap around $x = 0$. This applies whether the orbits are cantori, periodic orbits, or the upper and lower minimising orbits homoclinic to periodic orbits. On the other hand, every minimax orbit appears to have a point on the line $x = 0$.

Thus it is natural to choose turnstiles that have their centrepoints on this line. Then the union of these turnstiles forms a "chimney": all vertical transport takes place in this region. Our numerical work for the standard map suggests strongly that the iterates of the largest gap cover the partial separatrix completely.

This structure is a result of the symmetry of the standard map. We expect it to generalise to the class of reversible maps, i.e. those possessing a "reflection" S (i.e. orientation reversing map with $S^2 =$ identity) which conjugates the map to its inverse ($STS = T^{-1}$). They always seem to have a similar, "dominant" symmetry line on which each minimax orbit has a point. Thus we can choose this point to be the centre of the corresponding turnstile. Then the turnstiles all line up along this

dominant line, and the resulting "chimney" spans this line.

We have no theory of chimneys which applies generally, but we believe that the idea can be extended with care to more difficult cases in which the structure of the turnstiles may be more complicated, or there may be more than one independent turnstile in a cantorus or partial separatrix.

Since the stable manifolds of different periodic orbits cannot intersect (nor can the unstable manifolds), the turnstiles are expected to line up in a nice fashion as sketched in fig. 9.

3. Areas and actions

Areas of resonances and of turnstiles are both needed for the theory of transport. An obvious way to calculate them is to approximate the boundaries by closely spaced points and then use numerical integration; however, this is not the best way. In [1], we showed that the area of the turnstiles can be obtained from the difference in action between minimising and minimax homoclinic orbits. We now show how the areas of resonances can also be obtained from the actions of homoclinic orbits.

3.1. *Fundamental formula*

The basic formula relating action to area is illustrated in fig. 4. Let $\mathscr{C}$ be a directed curve in the phase plane. Parametrise it by λ ranging over $[0,1]$, so that

$$\mathscr{C}(\lambda) = \{x(\lambda), p(\lambda)\}. \tag{3.1}$$

We define the *algebraic area*, A, "under" $\mathscr{C}$ to be the signed area bounded by the loop formed from $\mathscr{C}$, the vertical lines $x = x(0)$ and $x = x(1)$ and the horizontal line $p = 0$. The direction of the loop is set as that of increasing λ along $\mathscr{C}$. Regions encircled by (counter-) clockwise loops are defined to have (negative) positive area, as usual. For the simple situation depicted in fig. 4, A is merely the geometric area. If, however, $\mathscr{C}$ intersects itself or if $p(\lambda)$ is negative for some range of λ, then the sign of the areas of these regions will change, and A will not be the geometric area under $\mathscr{C}$. In any case we will still refer to A as the area "under" $\mathscr{C}$, though some regions may be included with negative sign.

The image of $\mathscr{C}$ under the map T is $T(\mathscr{C}) = \mathscr{C}'$. The area under $\mathscr{C}'$ is denoted A', and, as is shown in fig. 4, is the signed area enclosed by $\mathscr{C}'$, the verticals $x = x'(0)$ and $x = x'(1)$, and the horizontal $p = 0$.

Let $F(x, x')$ be the generating function of the twist map T from the initial point with angle $x(\lambda)$ to its image point with angle $x'(\lambda)$. If the corresponding momenta are defined by (2.3), then

$$\frac{dF(x,x')}{d\lambda} = \frac{\partial F(x,x')}{\partial x'}\frac{dx'}{d\lambda} + \frac{\partial F(x,x')}{\partial x}\frac{dx}{d\lambda} = p'\frac{dx'}{d\lambda} - p\frac{dx}{d\lambda}. \tag{3.2}$$

Integrating both sides with respect to λ, we obtain

$$\Delta F \equiv F[x(1), x'(1)] - F[x(0), x'(0)] = \int_0^1 p'\frac{dx'}{d\lambda}\,d\lambda - \int_0^1 p\frac{dx}{d\lambda}\,d\lambda. \tag{3.3}$$

It is easy to see that the integrals in (3.3) actually represent the algebraic area, as we have defined it above, thus

$$\Delta F = A' - A. \tag{3.4}$$

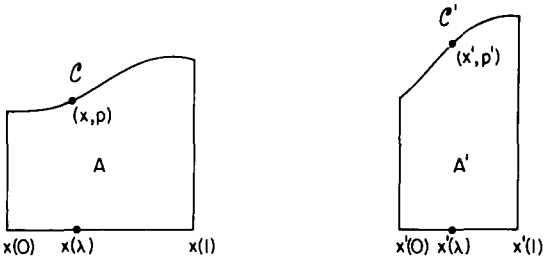

Fig. 4. Illustration of the relation between action and areas, eq. (3.4).

Eq. (3.4) is the basic formula from which all the others follow.

3.2 *Stable and unstable segments*

Two points Y_0 and Z_0 are called *future asymptotic* if they are distinct, but their orbits approach each other asymptotically, so as to become indistinguishable at sufficiently long times in the future:

$$\lim_{t\to\infty} |Y_t - Z_t| = 0, \tag{3.5}$$

where $|\ |$ represents any norm. Similarly they are *past asymptotic* if they are distinct and their orbits approach each other asymptotically in the past:

$$\lim_{t\to-\infty} |Y_t - Z_t| = 0. \tag{3.6}$$

Points which are both future and past asymptotic are *homoclinic* (to each other). If Y_0 is past asymptotic to W_0 and future asymptotic to Z_0 then it is *heteroclinic* from W_0 to Z_0.

If an orbit X_t is hyperbolic (the notion of hyperbolicity is not limited to periodic orbits; see Lanford [12] for an introduction), then the set of points which are future or past asymptotic to X_0 form two smooth curves without self-intersection, crossing transversely at X_0, called the stable and unstable manifolds of X_0. All points on the same stable manifold are future asymptotic, and all points on the same unstable manifold are past asymptotic. Given two such points we call the piece of invariant manifold between them a *stable* or *unstable segment*. The areas under such segments are important for the theory of transport, and will lead to formulae for the area of resonances.

We can find stable (unstable) segments numerically by taking the limit of backward (forward) iterates of straight lines joining corresponding points of two future (past) asymptotic orbits. Thus if Y_0 and Z_0 are future asymptotic, let $\mathscr{L}_j$, $j>0$, be the directed straight line segment from Y_j to Z_j.

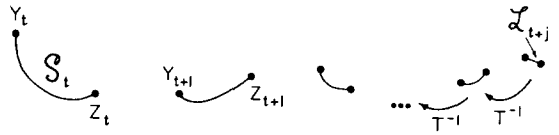

Fig. 5. Stable segments, $\mathscr{S}_t$, for a future asymptotic orbit pair $\{Y_t\}$, $\{Z_t\}$ are obtained from the preimages of the line segment $\mathscr{L}_{t+j}$ as j approaches ∞.

Then the stable segment joining Y_0 to Z_0 is

$$\mathscr{S}_0 = \lim_{j\to\infty} T^{-j}(\mathscr{L}_j). \tag{3.7}$$

This construction is illustrated in fig. 5. Similarly a pair of past asymptotic points gives an unstable segment

$$\mathscr{U}_0 = \lim_{j\to\infty} T^{j}(\mathscr{L}_{-j}). \tag{3.8}$$

There is no particular significance in choosing $\mathscr{L}_j$ to be straight lines: at least in the case of rational rotation number, any Lipschitz graphs over the stable manifold (in stable–unstable coordinates near the hyperbolic points) with uniform Lipschitz constant would give the same segments $\mathscr{S}_0$ and $\mathscr{U}_0$. The images of a stable (unstable) segment are also stable (unstable) segments, and are denoted $\mathscr{S}_t(\mathscr{U}_t)$.

Using the fundamental formula (3.4), the area below a stable or unstable segment can be expressed in terms of sums of action differences. Let $\{Y_t\}$ and $\{Z_t\}$ be a future asymptotic pair. Denote the angle coordinate of Y_t by y_t and of Z_t by z_t and the action difference by

$$\Delta F_t \equiv F(z_t, z_{t+1}) - F(y_t, y_{t+1}). \tag{3.9}$$

The future action difference sum is defined as

$$\Delta W_t^{\mathrm{f}} \equiv \sum_{j=t}^{\infty} \Delta F_j. \tag{3.10}$$

Similarly if $\{Y_t\}$ and $\{Z_t\}$ are past asymptotic

orbits then the past action difference sum is

$$\Delta W_t^{\mathrm{p}} \equiv \sum_{j=-\infty}^{t-1} \Delta F_j, \tag{3.11}$$

where ΔF is still given by (3.9). Note that ΔW_t^{f} includes ΔF_t while ΔW_t^{p} does not. The convergence of these sums is guaranteed if the union of the two orbits, plus their translates in angle by integers, is monotone. This is always the case whenever Y_t and Z_t are minimising orbits with the same "rotation symbol" m/n or $m/n+$ or $m/n-$ or ω (irrational), or one is minimising and the other is an associated minimax orbit because $\Sigma|z_t - y_t| \leq 1$.

Parametrise the stable segment $\mathscr{S}_t$ of a future asymptotic pair $\{Y_t\}$ and $\{Z_t\}$ by λ, such that $\mathscr{S}_t(0) = Y_t$, $\mathscr{S}_t(1) = Z_t$ and $\mathscr{S}_{t+1}(\lambda) = T\mathscr{S}_t(\lambda)$. The area under $\mathscr{S}_t$, denoted A_t^{s}, is obtained by iterating eq. (3.4):

$$\begin{aligned} A_t^{\mathrm{s}} &= A_{t+1}^{\mathrm{s}} - \Delta F_t = A_{t+2}^{\mathrm{s}} - \Delta F_{t+1} - \Delta F_t = \ldots \\ &= A_{t+k}^{\mathrm{s}} - \sum_{j=0}^{k-1} \Delta F_{j+t} \\ &= -\sum_{j=t}^{\infty} \Delta F_j \left(\text{since } \lim_{k\to\infty} A_k^{\mathrm{s}} = 0\right) \\ &= -\Delta W_t^{\mathrm{f}}. \end{aligned} \tag{3.12}$$

If $\{Y_t\}$ and $\{Z_t\}$ are past asymptotic, and A_t^{u} is the area under their unstable segment $\mathscr{U}_t$, then a similar calculation gives

$$A_t^{\mathrm{u}} = \Delta W_t^{\mathrm{p}}. \tag{3.13}$$

Note that the sign in (3.13) is indeed different from that in (3.12).

We can combine (3.12) and (3.13) if $\{Y_t\}$ and $\{Z_t\}$ are homoclinic. The signed area between the unstable and stable segments, which is positive where $\mathscr{U}_t$ is above $\mathscr{S}_t$, is given by

$$A_t^{\mathrm{u}} - A_t^{\mathrm{s}} = \Delta W \equiv \sum_{j=-\infty}^{\infty} \Delta F_j, \tag{3.14}$$

which is the difference in actions of the two orbits. Note that it is independent of t. This is because T is an area-preserving map, and hence the region contained between the stable and unstable segments has the same area for all time.

Once the area under single segment is known, we obtain the total area under all the segments $\mathscr{S}_t$ after time t by simply summing:

$$A_t^{\mathrm{f}} \equiv \sum_{k=1}^{\infty} A_{t+k}^{\mathrm{s}} = -\sum_{k=1}^{\infty} \Delta W_{t+k}^{\mathrm{f}} = -\sum_{k=1}^{\infty} k\Delta F_{t+k}. \tag{3.15}$$

Similarly the total area under all the unstable segments at time t and before is

$$\begin{aligned} A_t^{\mathrm{p}} &\equiv \sum_{k=-\infty}^{0} A_{t+k}^{\mathrm{u}} = \sum_{k=-\infty}^{0} \Delta W_{t+k}^{\mathrm{p}} \\ &= -\sum_{k=-\infty}^{0} k\Delta F_{t+k}. \end{aligned} \tag{3.16}$$

Finally, if $\{Y_t\}$ and $\{Z_t\}$ are homoclinic, then the total area under all the future stable segments and the past unstable segments (with the choice that the unstable segment is used at time t) is

$$A_t^{\mathrm{f}} + A_t^{\mathrm{p}} = A_t \equiv -\sum_{k=-\infty}^{\infty} k\Delta F_{t+k}. \tag{3.17}$$

In general, eq. (3.17) is not independent of the time t (where the shift from unstable to stable segment is made) and does not give a useful quantity. In fact we see that $A_{t+1} = A_t + \Delta W$, where ΔW is given by (3.14). There are special cases, however, for which ΔW is zero, and then (3.17) is independent of the origin of time. This brings us to the applications of these results.

3.3. *Area under partial separatrices*

Now we use eq. (3.17) to obtain the area under an upper partial separatrix for the simplest case of a 0/1 resonance around an orbit of period 1, or fixed point $x = x_{\mathrm{F}} = x_\ell = x_{\mathrm{r}}$, as illustrated in fig.

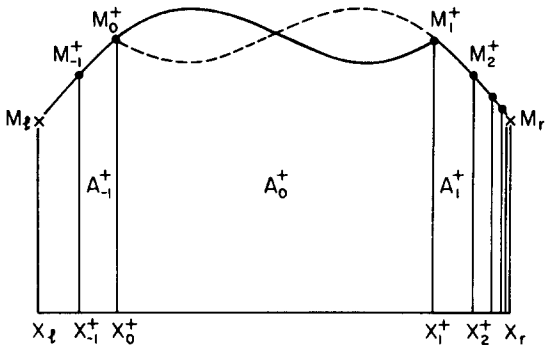

Fig. 6. Area under the upper partial separatrix of a fixed point, eq. (3.19). M_ℓ and M_r (the same point) represent the minimising fixed point.

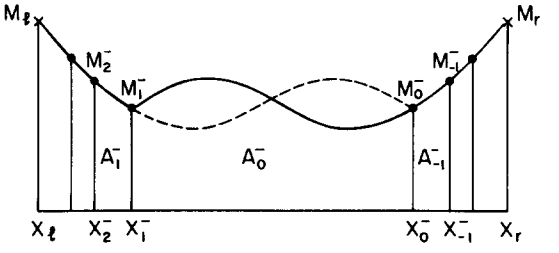

Fig. 7. Area under the lower partial separatrix of a fixed point, eq. (3.20).

6. For this simple case we choose Y_t to be the point M_t^+ of an upper minimising homoclinic orbit and $Z_t = M_{t+1}^+$ to be the next point on the same orbit. The orbits $\{M_t^+\}$ and $\{M_{t+1}^+\}$ are homoclinic, since they are both homoclinic to the fixed point. Furthermore since they are the same orbit, the action difference, (3.14), vanishes. The area under the complete upper partial separatrix is therefore

$$A^+ = -\sum_{t=-\infty}^{\infty} t\left[F(x_t^+, x_{t+1}^+) - F(x_{t-1}^+, x_t^+)\right]. \tag{3.18}$$

Another formula for A^+ can be obtained from (3.18) by subtracting the action of the fixed point, $F(x_F, x_F)$, from the first term in the brackets in (3.18) while simultaneously adding it to the second. This is necessary to maintain convergence. Now shift t by one in the second term to obtain

$$A^+ = \sum_{t=-\infty}^{\infty} \left[F(x_t^+, x_{t+1}^+) - F(x_F, x_F)\right]. \tag{3.19}$$

Fig. 7 illustrates the lower partial separatrix. Notice that the points of the homoclinic orbit move to the left in this case, so that increasing t means decreasing x. Defining Z_t and Y_t as successive points on the lower minimising orbit we obtain

$$A^- = \sum_{t=-\infty}^{\infty} \left[F(x_F, x_F) - F(x_t^-, x_{t+1}^-)\right]. \tag{3.20}$$

The final result is that the area in the 0/1 resonance is

$$A = A^+ - A^-. \tag{3.21}$$

It may seem surprising that the contributions to the area of the resonance from the fixed point action $F(x_F, x_F)$ in (3.19) and (3.20) add together instead of cancelling, but this is so, and comes from the fact that the asymptotic motion approaches the periodic orbit from the left in the upper separatrix, and from the right in the lower separatrix.

The analysis for an arbitrary m/n resonance is similar, and the same figs. 6 and 7 can be used, though now x_ℓ and x_r represent two neighboring angle coordinates on the orbit. We choose $Y_t = M_t^+$ and $Z_t = M_{t+n}^+$. Again, since ΔW for these orbits is zero, we have the area under the complete upper partial separatrix as

$$A^+ = -\sum_{t=-\infty}^{\infty} t\left[F(x_{t+n}^+, x_{t+n+1}^+) - F(x_t^+, x_{t+1}^+)\right]. \tag{3.22}$$

This sum can be reordered, adding and subtracting the action of the periodic orbit, and t shifted by n in the last term to obtain

$$A^+ = n\sum_{i=-\infty}^{\infty}\sum_{t=1}^{n} \left[F(x_{in+t}^+, x_{in+t+1}^+) - F(x_t, x_{t+1})\right]. \tag{3.23}$$

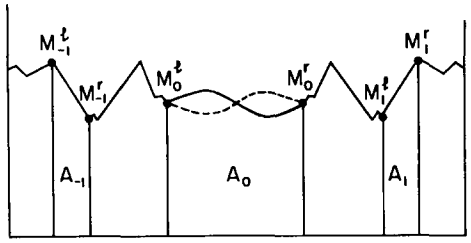

Fig. 8. Area under a cantorus partial barrier. The orbit of the left gap endpoint is denoted by superscript "ℓ" and of the right gap endpoint by "r".

The expression for A^- is identical except for signs.

In the above analysis, it has been assumed that there is only one minimising m/n orbit. If there is more than one such orbit, then each gives a family of gaps and one has to sum the contributions from each family.

3.4. *Area under partial barriers formed from cantori*

Another result which follows from (3.17) is the area under the partial barrier formed from a cantorus, as is illustrated in fig. 8. In this case we let $\{Y_t\} = \{M_t^{\ell}\}$ be the orbit of the left endpoints of a gap in the cantorus and $\{Z_t\} = \{M_t^{r}\}$ be the orbit of the right endpoints. These orbits are absolute minima of the action, and the action difference ΔW, (3.14), vanishes as proved by Mather [13]. Backward iterates of the unstable segment of a gap and forward iterates of the stable segment form the cantorus partial barrier. Eq. (3.17) directly gives the area under the gaps of the cantorus.

The total area under the partial barrier is thus given by the sum of (3.17) and the area under the cantorus itself. Numerical work, however indicates that cantori have zero length [14], and therefore that the area below them is zero. Indeed this has been proved under the hypothesis of non-zero "phonon gap" [15]. In appendix 2 we show how to modify this proof so that it applies in the dynamically more significant case of hyperbolicity. Thus in the hyperbolic case the area under the partial barrier is given by (3.17) alone.

Table I
Areas and actions. The sums are over action differences evaluated on the homoclinic pair of orbits $\{Y_t\}$, $\{Z_t\}$, defined in (3.9). M denotes a minimising orbit, S a minimax orbit, and H a heteroclinic orbit. Superscripts $\pm$ denote orbits homoclinic to the periodic orbit from the left (right), and ℓ and r denote orbits of the left or right gap endpoints of the cantorus, or the left and right heteroclinic orbits

	Area $= -\sum_{t=-\infty}^{\infty} t\Delta F_t$		Flux $= \sum_{t=-\infty}^{\infty} \Delta F_t$	
Orbit	Y_t	Z_t	Y_t	Z_t
Periodic m/n			M_t	S_t
Upper partial separatrix of m/n orbit	M_t^+	M_{t+n}^+	M_t^+	S_t^+
Lower partial separatrix of m/n orbit	M_t^-	M_{t-n}^-	M_t^-	S_t^-
Cantorus	M_t^{ℓ}	M_t^{r}	M_t^{ℓ}	S_t
Heteroclinic from m/n to m'/n'			H_t^{ℓ}	H_t^{r}

Note that the area under a partial barrier is independent of the construction of the partial barrier itself, depending only on the orbit of the gap endpoints. This is true for the separatrices as well. These formulae are summarized in table I.

Finally, we admit that we have no formula for the area under a KAM surface (invariant circle) in terms of the actions of a finite number of orbits. One might have thought this would be the simplest case!

3.5. *Flux*

Eq. (3.14) defines the flux through a homoclinic pair of orbits. For example we can let $\{Y_t\}$ be a minimising orbit $\{M_t\}$ and $\{Z_t\}$ be the corresponding minimax orbit $\{S_t\}$, corresponding to a cantorus or a partial separatrix. In this case (3.14) implies that the upward flux flowing between $\{M_t\}$ and $\{S_t\}$ is the difference in actions of these orbits.

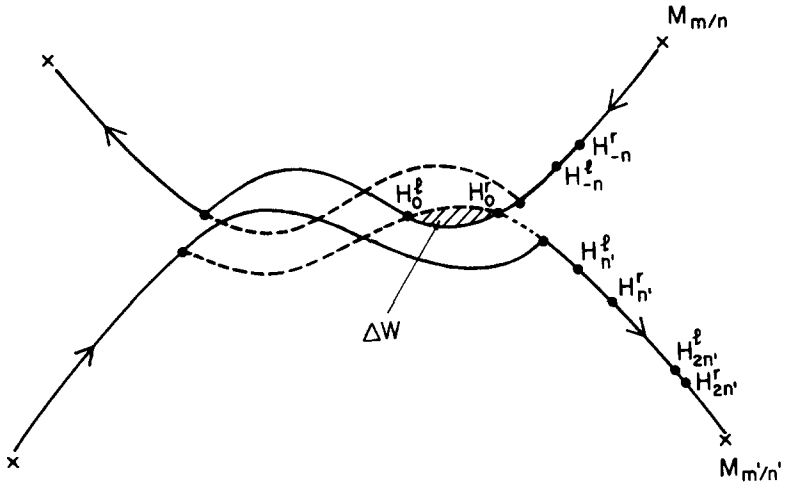

Fig. 9. Sketch of the area contained in the intersection of two turnstiles. The orbits H^{ℓ} and H^{r} are heteroclinic, approaching the upper resonance minimising orbit as $t \to -\infty$ and the lower resonance minimising orbit as $t \to \infty$.

For a resonance, a complete partial barrier can be constructed from the collection of all the past unstable and future stable segments connecting M_t to S_t, and the past unstable and future stable segments connecting S_t to M_{t+n} (see fig. 1). The downward flux is the difference in action between the latter two orbits, which is of course the negative of the upward flux. Analogously a partial barrier for a cantorus is constructed from the segments connecting M_t^l to S_t and those connecting S_t to M_t^r. If the iterates of the chosen gap fill in all the gaps of the cantorus, then ΔW is the total flux through the cantorus. Furthermore, since the action of M_t^{ℓ} and M_t^r are the same, the net flux is zero.

The upper turnstile area of an m/n resonance gives the area that makes a transition from inside the m/n resonance to some resonance above m/n. Similarly the lower turnstile represents the area making a transition to below m/n.

For calculations of transport it is not sufficient to know only these two areas; what we need to know is to which resonance the transition occurs, or equivalently how much of the m/n turnstile area is in a given resonance m'/n'. This can again be related to a difference in action, this time between orbits heteroclinic from the m/n minimising orbit to the m'/n' minimising orbit, as depicted in fig. 9. There are two such orbits in the right lobe of the turnstile, labelled H^{ℓ} and H^{r}, which are homoclinic to each other. Eq. (3.13) then implies that the difference in action between these two orbits is the area of the lower m/n turnstile that is below the m'/n' partial separatrix. This gives the transition flux $\Delta W_{m/n \to m'/n'}$. The flux $\Delta W_{m'/n' \to m/n}$ is given by the differences in action between the left lobe heteroclinic orbits.

These formulae are also summarized in table I.

4. Numerical results

4.1. *Finding periodic orbits*

It is a delicate matter to devise an algorithm to find highly unstable periodic orbits. The most straightforward method is based on looking for a zero of a function dependent on the nth iterate of the map. This is bound to fail when the product of the largest eigenvalue of the linearization of the map with the smallest change in initial condition is comparable to or exceeds the range of monotonicity of the function. Alternative methods, based

on minimisation of the action, are actually more stable when the eigenvalue is large, indeed the rate of convergence turns out to be proportional to this eigenvalue.

For our purposes, we found it sufficient to use a variant of Aubry's gradient method to find periodic orbits [2]. This technique involves solving the set of n equations

$$\mathrm{d}x_t/\mathrm{d}\tau = -\partial W/\partial x_t, \quad t = 0,1,2,\ldots,n-1 \tag{4.1}$$

in a fictitious time, τ. A point, $x_0,\ldots,x_{n-1}$, with $\nabla W = 0$, is an orbit, and is an attracting fixed point of (4.1) in τ if W is a minimum. To obtain the minimising orbit of frequency m/n, one merely solves (4.1) from some initial condition $x_t(0)$, imposing the constraint $x_n = x_0 + m$. Choice of the initial condition is important, however, since the function W can have many local minima besides the absolute minima. Note also that there are also lots of absolute minima corresponding to phase-shifting any one.

Following [16] we use the initial condition $x_t(0) = mt/n + \alpha$, where α is a phase whose value does not matter much. Angenent [17] showed that the gradient flow (4.1) has the property that if you start with an initial condition which is ordered like a rotation, it will remain so for all time. Since the desired orbits are so ordered this is a useful property. For reversible maps (the only ones we have considered in practice), we can obtain the minimax orbit by fixing $x_0(\tau)$ to be on the dominant symmetry line of the map (e.g. for (2.3), the line $x = 0$). The flow (4.1) restricted to this symmetry line has as an attracting fixed point the minimax orbit. One can also use symmetry to halve the number of points required. After finding the saddle, we shift the points on this orbit by a small amount, allowing the further relaxation into the minimising orbit.

Recently, it has been suggested that highly unstable orbits can be found more efficiently by using an initial condition which minimises W in the limit of large k, i.e. $x_t(0)$ are chosen at the maxima of $V(x)$ (2.3), and also ordered properly [18]. For further ideas see [28, 29].

To solve (4.1) we use a semi-implicit method, which involves inverting the linearization of the right-hand side to allow larger time steps. This technique limits to Newton's method as the time step is taken arbitrarily large.

The rate of convergence of the approximate orbit to the zero of ∇W is related to the eigenvalues of the Hessian matrix of second variations of the action, $H_{jk} \equiv \partial^2 W/\partial x_j \partial x_k$. We have shown previously [19] that the determinant of this matrix evaluated on an orbit is related to the residue as defined by Greene [20],

$$R \equiv \tfrac{1}{4}[2 - \mathrm{Tr}(DT^n)]. \tag{4.2}$$

For the case of $F_{12} = -1$, as in (2.2), we find $\det(\mathsf{H}) = -4R$. For a minimising orbit, the second variation of the action must be non-negative, so it has negative residue and is therefore hyperbolic. If the orbit is very unstable, then some of the eigenvalues of H must be quite large (say $\mathcal{O}(R^{1/n})$). This implies that, providing the initial condition for (4.1) is close enough to the minimising orbit for the linearized matrix to be relevant, convergence is rapid. However, there is typically at least one small eigenvalue corresponding to the direction of the minimax orbit, and this slows the convergence. Also one expects that the region of validity for linearization actually shrinks as the residue increases, and so one runs the risk of falling out of the basin of the minimising orbit.

4.2. *Finding homoclinic orbits*

Computation of the area of a resonance requires finding the homoclinic orbits to the minimising periodic orbit. Suppose we have a resonance of frequency m/n, with corresponding minimising periodic orbit M, and saddle S. The minimising and minimax homoclinic orbits are denoted M^+ and S^+ for the upper separatrix, and M^- and S^- for the lower separatrix.

Let m/n have a continued fraction expansion $[a_0, a_1, \ldots, a_l]$. The convergents to m/n are the truncations of the continued fraction at some level less than 1. If m_u/n_u is the convergent closest to m/n such that $m_u/n_u > m/n$, then the following sequence of frequencies converges to m/n from above as $j \to \infty$:

$$m_j/n_j = (m_u + jm)/(n_u + jn), \quad j = 1, 2, \ldots . \tag{4.3}$$

There are two periodic orbits for each j: the minimising and minimax orbits. As $j \to \infty$ these limit to the corresponding minimising and minimax homoclinic orbits, M^+, and S^+. As j increases, the number of points in the neighborhood of M continually increases as well, but the number of points over the remainder of the separatrix quickly stabilizes. The residue, R, of the m_j/n_j orbit is used as a criterion for convergence of the orbit to the separatrix. As j gets large this residue tends to grow exponentially with j; in fact, it should grow roughly as $R \approx (-R_M)^j$ since all but a finite number of points on the homoclinic orbit are arbitrarily close to M, and so their linear behavior is the same as that of M. Similarly the error in the resonance area tends to decrease exponentially at the rate $(-R_M)^{-j}$. Thus the error due to truncation of the sum at a period $n_j \approx j_n$ will be of size $\mathcal{O}(R^{-1})$. We pick j so that $R \approx 10^{10}$. For example, with $k \approx \mathcal{O}(1)$, the orbits homoclinic to $1/3$ are well approximated with $j = 14$. It is harder to approximate the separatrix of a resonance with residue of order one.

An easier way to find the homoclinic orbits when the residue is small is to search for intersections of the stable and unstable manifolds of M. For reversible maps one merely searches for the crossing of the dominant symmetry line by the unstable manifold to find the minimax homoclinic orbit. Begin by obtaining the eigenvectors of M from the linearization of the map. Put a point on the unstable eigenvector very close to a point on M. Iterate forward in time. Choose the number of iterations so that the point crosses the symmetry line (this number can be estimated from the eigenvalue at the fixed point). A zero finding routine can be used to adjust the initial position until its iterate falls on the symmetry line. Another way to find the orbits required in this paper is to use the fact that they are ordered. One can find them by iterating an initial point until its orbit gets out of order and then pushing the initial condition appropriately to get it back in order, e.g. using a bisection method. This is how some of the pictures of cantori in [1] were produced, and is the method used in [14].

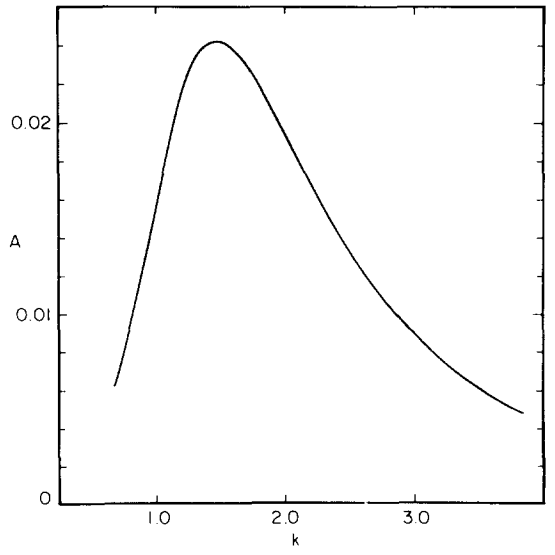

Fig. 10. Area of the 2/5 resonance for the standard map as a function of the parameter k. The area reaches a maximum of 0.0242 at $k = 1.45$. At this point the residue of the minimising orbit is -2.08, and the minimax orbit is unstable. In fact the period doubling sequence of the minimax orbit accumulates near $k = 1.42$.

4.3. *Area of a resonance*

The area of a single resonance, $A_{m/n} = A^+ - A^-$, as a function of the parameter for the standard map is shown in fig. 10. For small k, area is increasing as one would expect from perturbation theory (a resonance with a frequency m/n appears at the nth order in perturbation theory and its size should grow as $k^{n/2}$). As k increases the resonance begins to look less like the standard resonance of the pendulum Hamiltonian for which the estimate width and area $\propto$ (resonant

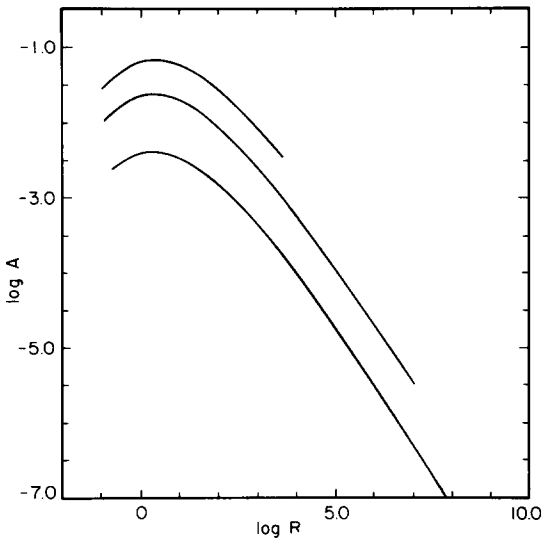

Fig. 11. Area of the 1/3, 2/5, and 4/11 resonances of the standard map as a function of $|R|$ of the minimising orbit. The smaller period resonances have larger area. Note that all areas peak near $R = -2$, and for large residue they decrease as a power of $|R|$, where the power approaches that given in (4.4) for large period.

forcing)$^{1/2}$ holds. When the residue is $\mathcal{O}(1)$, its area reaches a maximum and begins to decrease as $k^{-\eta}$, where η depends on the resonance frequency. For large residue, the upper and lower turnstiles nearly coincide and a resonance island has a characteristic "s"-shape in the gap near the dominant symmetry. Elsewhere, the island shapes are more nearly rectangular; the sharp corners are due to the switch from stable to unstable manifolds.

The area of a typical resonance as a function of residue is shown in fig. 11. For large residue the area decreases geometrically as $|R|^{-\lambda}$, where λ depends on the frequency. If we plot the area of various m/n resonances at a some fixed parameter value as a function of residue, we find a universal law:

$$A \propto |R|^{-\lambda} \tag{4.4}$$

with $\lambda \approx 0.939(\pm 4)$ over the range $100 < |R| < 10^{10}$. This seems to hold for any parameter value. It also describes the area of a single resonance as a function of its residue, providing $n \gg 1$, and is a reasonable estimate even for small n; for example, the 3/7 area decreases as the -0.87 power of R, etc.

The 0/1 resonance of the standard map behaves somewhat differently. Its area monotonically increases, approaching the value $A_{0/1} = 1$ for large k (implying $\eta = \lambda = 0$). This is due to the reflection symmetry of the map about the origin which implies the upper and lower separatrix have the same shape. Furthermore, the periodicity in the momentum direction implies that as k increases, each of the resonances $m/1$ are equivalent. We suspect that if other driving harmonics were added to the standard map, these would also give resonances whose area did not go to zero for large k.

4.4. *Total resonance area*

To add the area of all the resonances, we use the Farey tree procedure to order the rationals. The Farey tree seems to pick out rationals in order of their importance. As was discussed in [1], the area of turnstiles strictly decreases as one moves down the Farey tree. Here we find that the area of resonances also decreases.

To obtain a Farey tree, define the zeroth generation by a pair of rationals, m_1/n_1, and m_2/n_2 satisfying $m_1 n_2 - m_1 n_2 = \pm 1$ (such a pair is called "neighboring"). The gth generation is constructed by adding numerators and denominators of each pair of neighboring rationals for all generations less than g. Thus the first generation consists of the rational $(m_1 + m_2)/(n_1 + n_2)$. Each rational has two daughter rationals, obtained by adding its numerator and denominator to its two neighbors. This forms a binary tree; there are 2^{g-1} rationals in the gth generation. As is well known, every rational in the interval $[m_1/n_1, m_2/n_2]$ occurs exactly once on the tree.

Consider, for example, the standard map at a parameter value above the critical value where the last rotational KAM surface is destroyed ($k_{cr} = 0.971635406$). Fig. 12 shows the areas A^+ and A^- for the homoclinic orbits for all rationals on the (1/3, 1/2) Farey tree up to the 7th generation, as a function of frequency. For example the 2/5

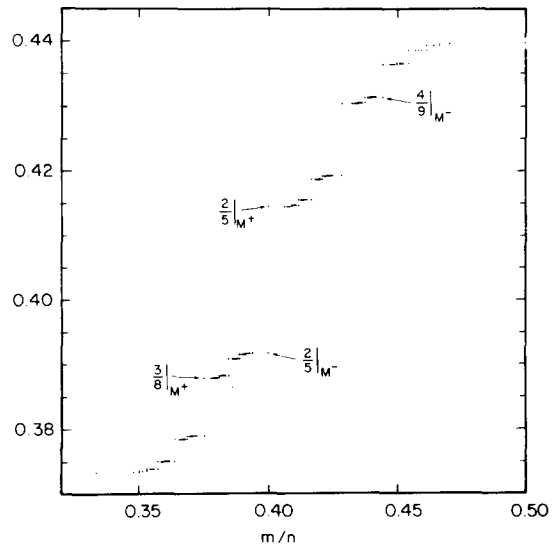

Fig. 12. Area below the upper and lower partial separatrices of periodic orbits of the standard map at $k = 1.283$. Periodic orbits are obtained from the Farey tree with starting rationals $[\frac{1}{3}, \frac{1}{2}]$ up to seven generations (129 + 2 periodic orbits). Residues range from -1 to $-4(10)^{11}$ (the latter is for the 34/81 orbit).

resonance has an area 0.3917164 below its lower separatrix, denoted as $2/5|_{M^-}$, and 0.4146119 below its upper separatrix, denoted as $2/5|_{M^+}$. The area of higher order rationals very quickly approaches zero, as indicated by the near coincidence of the upper and lower separatrix areas for larger generations in fig. 12.

The curve, "area as a function of frequency" can be thought of as a picture of the "action" function of a non-integrable system. This function is, by definition, non-decreasing but appears numerically to be almost everywhere flat. It has jumps across every rational according to the area of that resonance. From this picture it is clear why a perturbation series which attempts to find a smooth invariant must fail.

The total area of the region of phase space between the 1/3 and 1/2 resonances is given by the difference between the area below $1/2|_{M^-}$ and that below $1/3|_{M^+}$. This gives

$$A_{\text{tot}} = 6.63582(10)^{-2}. \tag{4.5}$$

The area of an individual resonance is obtained from the difference between the areas below the M^+ and M^- orbits of that resonance. The resonance areas on the Farey tree are shown in fig 13. Here we include every resonance whose mother has an area larger than 10^{-6}.

The area of all resonances between the $1/3|_{M^+}$ and $1/2|_{M^-}$ orbits is obtained by summing the area of each resonance on the Farey tree. Table II shows this sum for each generation through the seventh. To estimate the total area it is preferable not to sum in this fashion since, as seen in fig. 13, the decrease of area along Farey paths which limit to rationals is much slower than that along oscillating paths, so larger generations must be included in the former cases. Summing the areas of all the resonances shown in fig. 13 gives the value $A_{\text{res}} = 6.63581(10)^{-2}$, which is within 1 part in 10^{-6} of A_{tot}, (4.5).

We conclude that resonances occupy all of the phase space area. This calculation has been done for other values of the parameter k. As k increases, the area between the 1/3 and 1/2 resonances decreases, and more of the total phase space is taken up by the driven, $m/1$ resonances. However, so long as $k > k_{\text{cr}}$, we find that resonances fill phase space.

This implies that fig. 12 is a complete "devil's staircase": it is constant except for jumps at rational points. This is analogous to the devil's

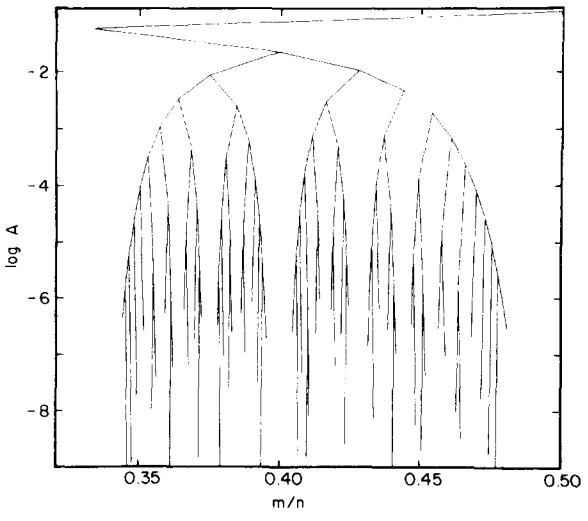

Fig. 13. Resonance areas on the $[\frac{1}{3}, \frac{1}{2}]$ Farey tree for the standard map at $k = 1.283$.

Table II
Area of resonances for Farey generations with starting rationals (1/3,1/2), at $k = 1.283$. Generation one corresponds to 2/5, etc.

Generation	Resonance area
1	0.22895E-01
2	0.19997E-01
3	0.14144E-01
4	0.66828E-02
5	0.19733E-02
6	0.48200E-03
7	0.12913E-03
Total	0.66304229E-01

staircase for circle maps at the critical parameter value [21, 22]. In this case a resonance area is represented by the width of a mode locking interval of frequency m/n.

Actually, Chen [30] has shown that the area under a partial barrier as a function of frequency ν is precisely Aubry's devil's staircase describing the lacking of epitaxial layers to a crystal [31], which is the derivative with respect to ν of the average Lagrangian

$$L_\nu = \lim \frac{1}{2n} \sum_{i=-n}^{n-1} h(x_i, x_{i+1})$$

for the corresponding orbits.

Note, however, that our Farey tree and devil's staircase are not geometrically self-similar. In fact the decrease of area along any path accelerates with increasing generation. This is due to the fact that for $k > k_{cr}$, no Farey paths approach frequencies which give invariant circles. Geometric scaling of areas should apply only for critical invariant circles. Indeed we have found previously that the scaling near a cantorus proceeds as the exponential of an exponential [1], and not geometrically, and we suspect that a similar result holds in this case.

4.5. *Partial barriers have measure zero*

Outside the resonances, there are rotational invariant circles and the partial barriers formed from cantori. The numerical results above indicate that the latter, as well as any other points that might exist outside the resonances have total measure zero. Cantori always seem to be hyperbolic. Indeed this can be proved [23] quite easily for the standard map when $k > 2\sqrt{\pi^2 + 1}$, and we need to assume hyperbolicity in order to define the partial barriers, at least with the techniques available at present. We will show that the union of the partial barriers for hyperbolic cantori have zero measure using Poincaré's recurrence theorem, as follows.

Let the sets $\mathscr{S}$ and $\mathscr{U}$ be the union of the pieces of stable/unstable manifold, respectively, used to close gaps in cantori with rotation numbers in some interval, excluding the endpoints. The forward orbit of every point in $\mathscr{S}$ converges to the orbit of some endpoint of some cantorus, so it is not forward recurrent. But $\mathscr{S}$ is contained in a bounded set, and $T\mathscr{S} \subset \mathscr{S}$, so by Poincaré's recurrence theorem $\mathscr{S}$ has measure zero. Similarly by considering T^{-1}, $\mathscr{U}$ has measure zero. The union of the cantori themselves has measure zero too since hyperbolic sets have non-zero Lyapunov exponents, and any invariant set with non-zero Lyapunov exponents has a decomposition into sets of positive measure which are ergodic [24] which would be impossible for a union of cantori [4]. Thus the union of the partial barriers formed from hyperbolic cantori has measure zero, as claimed.

5. Conclusions

Chaotic motion takes place within resonances, and a point remains within a given resonance until it reaches a turnstile, when it makes a transition to another resonance. Numerical results suggest that the resonances plus the rotational invariant circles form a partition of phase space, up to a set of measure zero. For $k > k_c$ there are no rotational invariant circles, and the total area is filled by resonances. For large enough k the area of all resonances, except the driven $(m/1)$ resonances, decreases. Thus, asymptotically for large k, phase space is filled entirely by the driven resonances.

The resonance picture of chaotic motion is complementary to the picture based on partial barriers formed from cantori. The motion is seen to take place from resonance to resonance: resonances could be the states in a coarse grained stochastic description. Resonances form a countable set, which the cantori do not. This could lead to an orbit coding scheme, where the code is a sequence of rationals designating which resonance an orbit is trapped in at which time. Our plan is to test numerically a Markov model based on resonances, and to look for universal features of the resonance partition to enable one to extrapolate from calculations of a small number of resonances.

For any two resonances m/n and m'/n', the flux from one to another is limited most strongly by the partial barrier with minimum flux [1]. The minimum flux barrier seems to be the cantorus with the most noble frequency between the resonance frequencies, e.g. $(m+\gamma m')/(n+\gamma n')$, when m/n and m'/n' are neighbours, ordered with $n' > n$, and γ is the golden mean $(1+\sqrt{5})/2$. The direct transition from m/n to m'/n' is possible only if the turnstiles of the resonances overlap. This implies that there is a maximum jump in frequency upon an iteration.

This provides a rigorous version of the resonance overlap criterion: if the upper turnstile for a resonance of rotation number m/n overlaps the lower turnstile for a resonance of rotation number $m'/n' > m/n$, then there are no rotational invariant circles with rotation number in the interval $(m/n, m'/n')$.

Below the value of k for the destruction of the last rotational invariant circle, one expects that the phase space of the divided system is filled by the two components, invariant curves and resonances, with the area of invariant curves going smoothly to zero as k approaches k_{cr}. We have not verified this in detail, since our technique for determining resonance areas by using periodic orbit approximations to the homoclinic orbits becomes more difficult as k decreases.

The geometry of the resonances is not simple. There are resonances corresponding to every rational, with the area monotonically decreasing as one moves down the Farey tree. For reversible maps, where the largest gaps line up on a symmetry line, there is a chimney in which all transport occurs. The range of possible transitions and the size of the chimney depend on the details of the area-preserving map, but far from the chimney a universal structure for the resonances is possible. The details remain to be worked out.

In order to use the theory directly for continuous time Hamiltonian systems of $1\frac{1}{2}$ or 2 degrees of freedom, our formulae need converting. We did this already for the formulae for flux [25], but conversion of the formulae for areas of resonances remains to be done.

For larger numbers of degrees of freedom there are in general no complete barriers to the chaotic motion. However it may still be possible to define resonances, and transitions between resonances, so the present picture opens up the possibility of treating transport in systems of arbitrary number of degrees of freedom on a common basis.

Acknowledgements

We would like to thank M. Kotschenreuther for allowing us to use his program for finding periodic orbits, and Q. Chen, A. Manning, J. Stark and P. Walters for useful discussions. This research was sponsored by a NATO grant, RG 85/0461, for international collaboration. One of us (JDM) was supported by US DOE grant DE-FG05-80ET-53088, and the others would like to thank the UK SERC for support.

Appendix 1

Partial separatrices and minimising semi-orbits

Katok [11] proved that given any point x_0 in a gap (l_0, r_0) in the set of points with minimising m/n_+ orbit or m/n orbit, there is a forward minimising orbit x_t^+, $t \geq 0$, with $x_0^+ = x_0$ and $x_t^+ \in (l_t, r_t)$, $t > 0$, and a backwards minimising orbit x_t^-, $t \leq 0$, with $x_0^- = x_0$ and $x_t^- \in (l_t, r_t)$, $t < 0$. These orbits are thus forward (backwards)

asymptotic to the orbits of the minimising m/n points containing (l_0, r_0) in the gap between them.

The only problem is that as x_0 moves from l_0 to r_0, the corresponding phase space point (x_0, p_0) need not move continuously. In particular, if a stable manifold has a turn in it so that it has two points with the same x-coordinate then by Aubry's Fundamental Lemma [2] they cannot both have minimising forward orbits. (A Maxwell rule determines which is minimising [26].) However, Stark [27] showed in the hyperbolic case that every point (x_0, y_0) of a stable manifold has an eventually minimising forward orbit, i.e. $\exists T$ such that the segments $[x_t, \ldots, x_{t'}]$ are minimising $\forall T < t < t'$. The points (x_0^+, y_0^+) do form a graph over x near hyperbolic minimising periodic points, thus providing an alternative construction of invariant manifolds.

One might have hoped, even though the invariant manifolds are usually not graphs over x, that one could choose a heteroclinic point at which to change over in order to make the partial barriers always be graphs. But this is not always possible. You can always modify a map by localised non-uniform rotations, preserving the twist property, to make the invariant manifolds turn over before they meet. As a corollary, one cannot always choose partial barriers so that every point has either minimising forward or minimising backward orbit.

Appendix 2

Hyperbolic cantori have zero length

Aubry et al. [15] in appendix H prove that if the "phonon gap" of a cantorus is non-zero, then the projection of the cantorus onto the angle coordinate has zero length. We show here how to extend their proof under the more dynamically relevant hypothesis of uniform hyperbolicity or, slightly weaker, positive Lyapunov exponent.

We use their notation. Here Q_n and P_n represent the configurations of two independent tangent orbits to the cantorus obtained from the initial conditions $P_0 = 0$, $P_1 = 1$, and $Q_0 = -1$, $Q_1 = 0$; $\Lambda_1 \equiv \lim Q_n/P_n$ as $n \to \infty$, and $\Lambda_2 \equiv \lim Q_n/P_n$ as $n \to -\infty$. The b_n are given by $-\partial^2 F/\partial x_n \partial x_{n+1}$ evaluated on a cantorus orbit; $b_n = 1$ for the standard map.

Lemma. There do not exist C, $\gamma > 0$ such that $\forall n > 0, Q_n - P_n\Lambda_1 \geq C\,\mathrm{e}^{n\gamma}$.

Proof. Suppose $\forall n > 0, Q_n - P_n\Lambda_1 \geq C\,\mathrm{e}^{n\gamma}$. Now $(Q_n - P_n\Lambda_1)/P_n \to 0$ as $n \to \infty$. So

$$P_n \geq C/K\,\mathrm{e}^{n\gamma},$$

for some $K > 0$. Eq. (H.12) of [15] implies that

$$Q_n/P_n - \Lambda_1 = \sum_{m \geq n} b_0/(b_m P_m P_{m+1})$$

$$\leq \sum_{m \geq n} b_0/b_m (K/C)^2\,\mathrm{e}^{-2m\gamma} \leq D\,\mathrm{e}^{-2n\gamma}.$$

This implies

$$P_n = (Q_n - P_n\Lambda_1)/(Q_n/P_n - \Lambda_1) \geq C/D\,\mathrm{e}^{3n\gamma}.$$

Repeating this, we get

$$P_n \geq C/D_m \exp(s_m n\gamma),$$

where $s_{m+1} = 2s_m + 1 \to \infty$. But $\exists C', \Gamma$ such that $P_n \leq C'\,\mathrm{e}^{n\Gamma}$, $n > 0$. This is a contradiction. Similarly, one can see that $Q_n - P_n\Lambda_2$ does not grow exponentially as $n \to -\infty$.

Remark. Actually we suspect that $Q_n - P_n\Lambda_1$ is always bounded as $n \to \infty$.

Proposition. Every uniformly hyperbolic cantorus, and cantorus with positive Lyapunov exponent, has zero projected length.

Proof. Suppose not. Then Aubry et al. prove existence of an orbit of the cantorus for which $\Lambda_1 = \Lambda_2$. This implies $h_n = Q_n - P_n\Lambda_1$ is a tangent orbit which does not grow exponentially in either direction of time. This contradicts hyperbolicity of

the cantorus, since every tangent orbit must grow exponentially in at least one direction of time.

If the hypothesis is weakened to positive Lyapunov exponent, then we use the fact that the set of points that Aubry et al. find with $\Lambda_1 = \Lambda_2$ has positive measure, with respect to the pullback of Haar measure under the semi-conjugacy to rotation, if the cantorus has non-zero length. Almost every point of the cantorus, with respect to this measure, has a Lyapunov exponent and it is the same almost everywhere. The Lyapunov exponent λ of the cantorus is defined to be this value. Thus there exists a point x with this value of exponent and $\Lambda_1 = \Lambda_2$. Since x has a tangent orbit which does not grow exponentially in either direction of time, we deduce that $\lambda = 0$.

Remark. Actually, we suspect that since cantori are uniquely ergodic, every orbit has the same exponent, in which case the above argument could be simplified.

Note added in proof

Actually there is a much easier proof that cantori with non-zero exponents have zero length, based on a general result of Young. In fact it shows the stronger result that they have dimension zero [32].

References

[1] R.S. MacKay, J.D. Meiss and I.C. Percival, Phys. Rev. Lett. 52 (1984) 697; Physica 13D (1984) 55.
[2] S. Aubry and P.Y. Le Daeron, Physica 8D (1983) 381.
[3] J.N. Mather, Topology 21 (1982) 457.
[4] A. Katok, Ergodic Theory and Dyn. Sys. 2 (1982) 185.
[5] B.V. Chirikov, Phys. Rep. 52 (1979) 265.
[6] A.N. Lichtenberg and M.V. Lieberman, Regular and Irregular Motion, (Springer, New York, 1982).
[7] S.R. Channon and J.L. Lebowitz, Ann. NY Acad. Sci. 357 (1980) 108.
[8] J.H. Bartlett, Cel. Mech. 28 (1982) 295.
[9] G.D. Birkhoff, Mem. Pont. Acad. Sci. Novi Lyncaei 1 (1935) 85, ch. 2, sect. 8.
[10] J.D. Meiss, Phys. Rev. A 34 (1986) 2375.
[11] A. Katok, More about Birkhoff periodic orbits and Mather sets for twist maps, IHES preprint (1982).
[12] O.E. Lanford, in: Chaotic behaviour of deterministic systems, Les Houches 1981, G. Iooss, R.H.G. Helleman and R. Stora, eds. (North Holland, Amsterdam, 1982).
[13] J.N. Mather, Publ. Math. IHES 63 (1986) 153.
[14] W. Li and P. Bak, Phys. Rev. Lett. 57 (1986) 655.
[15] S. Aubry, P.Y. LeDaeron and A. André, Classical ground-states of a one-dimensional model for incommensurate structures, Saclay preprint (1982).
[16] M. Peyrard and S. Aubry, J. Phys. C16 (1983) 1593.
[17] S.B. Angenent, The periodic orbits of an area-preserving twist map, Leiden preprint (1984).
[18] H.J. Schellnhuber, H. Urbshat and A. Block, Phys. Rev. A 33 (1986) 2856.
[19] R.S. MacKay and J.D. Meiss, Phys. Lett. A 98 (1983) 92.
[20] J.M. Greene, J. Math. Phys. 20 (1979) 1183.
[21] M.H. Jensen, P. Bak and T. Bohr, Phys. Rev. A 30 (1984) 1960.
[22] O.E. Lanford, Physica 14D (1985) 403.
[23] D. Goroff, Ergodic Theory and Dyn. Sys. 5 (1985) 337.
[24] Ya. B. Pesin, Sov. Math. Dok. 17 (1976) 196.
[25] R.S. MacKay and J.D. Meiss, J. Phys. 19 A (1986) L225.
[26] R. Graham and T. Tel, Phys. Rev. A 31 (1985) 1109.
[27] J. Stark, private communication (1985).
[28] B. Mestel and I.C. Percival, Physica 24D (1987) 172.
[29] Q. Chen, J.D. Meiss and I.C. Percival, Orbit extension method for finding unstable orbits, submitted to Physica D.
[30] Q. Chen, Devil's staircase in area-preserving maps, preprint, Austin TX.
[31] S. Aubry, Springer Lecture Notes in Math. 925 (1982) 221.
[32] R.S. MacKay, Hyperbolic cantori have dimension zero, J. Phys. A, to appear.